

# Dealing with Polysemy in Russian National Corpus: The Case of Adjectives

Ekaterina V. Rakhilina<sup>1</sup>, Tatiana I. Reznikova<sup>2</sup>, and Olga Yu. Shemanaeva<sup>3</sup>

<sup>1</sup>Institute of Russian Language, RAS, Moscow

<sup>2</sup>Dept. of Linguistic Studies, VINITI, RAS, Moscow

<sup>3</sup>Institute of Linguistics, Russian State University for Humanities, Moscow  
rakhilina@gmail.com, tanja\_reznikova@mail.ru,  
shemanaeva@yandex.ru

**Abstract.** The paper describes a research carried out in the Russian National Corpus project ([www.ruscorpora.ru](http://www.ruscorpora.ru)). We discuss a method of word sense disambiguation, which is now being applied to polysemous adjectives in the RNC. The approach implies formulating rules to select the appropriate sense of the adjective by using co-occurrence restrictions observed in the corpus. The disambiguating filters operate with various kinds of grammatical and semantic information on the adjectives and the nouns modified. Our results demonstrate that the semantic filters are effective for WSD.

**Keywords:** Corpus linguistics, word sense disambiguation, semantic annotation, Construction Grammar, Russian.

## 1 Introduction

In our work we discuss a method for word sense disambiguation, which is now being applied to the annotation system of the Russian National Corpus. Unlike most of the existing strategies that rely on statistical and machine learning methods (cf. the overview in [1]), our technique combines statistics with a rule-based approach. Disambiguation rules are formulated based on the statistical analysis of co-occurrence restrictions that can be observed in the corpus data. We claim that this approach reveals important generalizations, which are of high relevance for theoretical linguistics.

This paper presents the domain of adjectives denoting physical characteristics (such as temperature, colour, size, form, time, speed, etc.) or human properties (cf. ‘courageous’, ‘intelligent’, ‘honest’, etc.). Dealing with these data we will discuss the application of our method and its perspectives.

The study is based on the theoretical framework of Construction Grammar (cf. [2], [3], [4]). The Construction Grammar theory assumes that constructions – i.e. conventionalized pairings of form and meaning – are the basic units of language. All the constituents in a construction are bound into the whole entity; they are co-dependent and influence one another, which means that the change of any constituent leads to the change in the meaning of the whole pattern. In regard of this, polysemy can be viewed not as an independent property of a lexeme, but instead as its ability to

be *coerced* by particular constructions into having other meanings, cf. the notion of “coercion” in the works of R. Jackendoff [5], J. Pustejovsky [6], B. Partee [7].

Within Construction Grammar, the practical task of word sense disambiguation takes the following form: given a polysemous word, we have to formulate constitutive properties for each construction that can have this word as its part. Since different constructions are associated with different meanings, these properties can be used for context identification and accordingly for sense determination<sup>1</sup>.

The structure of the paper is as follows: Section 2 outlines the principles of semantic annotation in RNC. Section 3 illustrates the functioning of the rule-based system. Finally, the theoretical relevance of the approach is discussed in Section 4.

## 2 Semantic Annotation in RNC

The present research is carried out within the larger Russian National Corpus (RNC) project ([www.ruscorpora.ru](http://www.ruscorpora.ru)). The RNC currently contains over 140 mln. words and provides different kinds of annotation, particularly POS-tagging (with information on parts of speech and morphological features)<sup>2</sup>. What is unique for a large corpus of this kind is its semantic annotation (cf. [10]).

Semantic tags in the RNC correspond to conceptual categories assigned on the basis of vocabulary classification. The principles of lexical classification are derived from the project “Lexicograph” (<http://www.lexicograph.ru>) supervised by E. Paducheva and E. Rakhilina. The project aims at establishing a comprehensive database on the lexical semantics of Russian (up to the present moment, the study has mainly been focused on verbs and object nouns, for the theoretical ideas behind the project, see [11], [12], [13]). For the needs of the Russian National Corpus, the classification was revised and extended to cover all content words.

The classification follows the multi-faceted principle: there are several parameters (some of them hierarchical) independent of one another. At present, six classifications are involved in the annotation:

**Category** (e.g. “concrete nouns”: *stol* ‘table’, *sneg* ‘snow’; “abstract nouns”: *ljubov* ‘love’, *žara* ‘heat’; “proper nouns”: *Moskva* ‘Moscow’, *Ivan*; “qualitative adjectives”: *tverdyj* ‘hard’; “relational adjectives”: *kamennyj* ‘stony, made of stone’; “possessive adjectives”: *papin* ‘father’s’; “invariable adjectives”: *bež* ‘beige’);

**Taxonomy** (e.g. “weapon”: *puška* ‘cannon’, *ruž'e* ‘gun’; “device”: *gradusnik* ‘thermometer’, *telefon* ‘phone’; “space&place”: *gorod* ‘town’, *pole* ‘field’; “perception”: *sluh* ‘hearing’, *vzgljad* ‘look’; “emotion”: *pečal'* ‘sorrow’, *udovol'stvie* ‘delight’; “physical properties:form”: *krivoj* ‘curved’, *kruglyj* ‘round’; “size:large”: *vysokij* ‘high, tall’, *dlinnyj* ‘long’; “behaviour”: *krivljatsja*: ‘to make faces’, *skandalit'* ‘to brawl’);

**Mereology** (e.g. “building parts”: *dver'* ‘door’, *arka* ‘arch’; “sets and aggregates”: *mebel'* ‘furniture’, *trava* ‘grass’; “quanta and portions of stuff”: *kusoček* ‘lump’, *volna* ‘wave’);

<sup>1</sup> For the interaction of corpus linguistics and Construction Grammar cf. [8].

<sup>2</sup> The RNC includes also a corpus of syntactically annotated texts; see [9].

**Topology** (e.g. “container”: *stakan* ‘glass’, *dom* ‘house’; “holes”: *okno* ‘window’, *pora* ‘pore’; “ropes”: *tsep* ‘chain’, *nitka* ‘thread’);

**Evaluation** (e.g. “positive”: *aromat* ‘odor’, *četskij* ‘precise’; “negative”: *man’jak* ‘maniac’, *preslovutyi* ‘notorious’);

**Derivational classes** (e.g. “diminutives”: *knižēčka* ‘little book’, “adjectives derived from nouns”: *sosnovyj* ‘piny’).

The term Category refers to prime lexical divisions that determine main semantic and morphological features. Nouns are divided into abstract, concrete, and proper, while adjectives have four classes of qualitative, relational, possessive, and invariable. Relational adjectives differ from qualitative adjectives in that they are not gradable and cannot form comparatives.

The system of taxonomic classes is rather elaborated. It includes size, distance, quantity, time, physical and human properties for adjectives; people, animals, plants, buildings, devices, stuff, texts, food and drinks for concrete nouns; first and last names, patronymic names and toponyms for proper nouns; classes of abstract nouns are inherited mainly from verb and adjective hierarchies and include movement, impact, speech, human properties, colour, temperature, diseases, sports, parameters, etc.

Mereological annotation is applied to concrete nouns only. It provides a distinction between parts of the body, parts of instruments, clothes and other things as well as quanta & portions of stuff and phases of processes. The feature of sets and aggregates are used for such words as ‘set’, ‘bunch’, ‘furniture’, ‘mankind’. Nouns like ‘animal’, ‘fruit’, ‘instrument’, ‘name’ that denote categories of the world belong to the “names of classes” group.

The notion of topological types was put forward by L. Talmy [14], who has demonstrated their significance for the understanding of linguistic structures that describe space and shape as well as undoubted cross-linguistic relevance of geometric features. Names of physical objects associated with such topological types as “horizontal spaces”, “containers”, “holes”, “juts”, “ropes”, etc. appear to be sensitive to space operators, such as adjectives of form and size, prepositions, verbs and nouns which refer to form, location, and motion.

Lexical meanings that have positive or negative connotations form two classes in the category of Evaluation. Finally, derivational classes include words in which semantic components are introduced by a certain prefix or suffix or words derived from other parts of speech and what is more, from a particular semantic class of a particular POS (e.g. nouns derived from verbs; adjectives derived from names of substance).

Each content word in the vocabulary is classified along with all applicable parameters, the results are stored in the semantic dictionary of the corpus. For instance, the words *nora* ‘burrow’ and *naselenie* ‘population’ have the following attributes:

*nora* ‘burrow’

category: “concrete”, taxonomy: “space”, toponymy: container

*naselenie* ‘population’

category: “concrete”, mereology: “aggregate of persons”, derivational class: “derived from verbs”.

In case of a polysemous word, attributes are defined separately for each sense, cf. for the word *tihij*:

- tihij*<sub>1</sub> ‘low (about sounds)’  
category: “qualitative”, taxonomy: “sound”;
- tihij*<sub>2</sub> ‘quiet’  
category: “qualitative”, taxonomy: “human property”;
- tihij*<sub>3</sub> ‘faint’  
category: “qualitative”, taxonomy: “degree:minimal”;
- tihij*<sub>4</sub> ‘slow’  
category: “qualitative”, taxonomy: “speed:minimal”.

The semantic dictionary is then applied to the corpus. During the annotation process each content word in the RNC is automatically assigned all the tags that it has in the semantic dictionary. This in particular means that polysemous words are not disambiguated, that is, for instance, each occurrence of the word *tihij* receives four tags from the taxonomic classification: “sound” / “human property” / “degree” / “speed”.

On the part of the user, this means that the search for semantic characteristics of words returns many irrelevant results, e.g. the query for adjectives of “form” yields, among others, word combinations like *tupoj čelovek* ‘stupid person’ and *tupaja bol’* ‘dull pain’, as one of the meanings of the adjective *tupoj* refers to form, cf. *tupoj ugol* ‘obtuse angle’, or *botinki s tupym noskom* ‘square-tipped shoes’.

Thus, our goal is to distinguish the different meanings of adjectives and to provide the users of the Russian National Corpus with the semantically disambiguated texts. As a result, they could easily use them without any inconveniences such as finding inappropriate homonyms alongside with the needed word.

In order to achieve this goal, i.e. to avoid the polysemy of adjectives in RNC, we formulate rules (filters), which assign the only meaning to the adjective in the corresponding construction. Once the filter has been applied, all meanings of the target adjective that are inappropriate for the construction are deleted. The disambiguated adjectives are marked with the following features: 1) SEM (for a tag set that characterizes the first meaning listed in the dictionary), 2) SEM2 (for tag sets associated with other meanings) and 3) SEMF (for tag set(s) of disambiguated meanings). Thus, the subsequent queries in the corpus may focus only on the first meanings of the words, or on the disambiguated meanings.

### 3 The Rule-Based Approach to Disambiguation

The disambiguating rules are formulated manually on the database of 2- and 3-word clusters with associated frequency, POS, and semantic tags. The filters operate with the following information about the target adjective and the neighbour noun:

#### Morphosyntactic information

- grammatical features of the adjective (“long” vs. “short” form<sup>3</sup>; case, number; comparative, superlative)
- grammatical features of the neighbour noun (animate vs. non-animate<sup>4</sup>, case, number)

<sup>3</sup> Russian adjectives may appear in two forms: the long form, which has case, number, gender features, and the short form, which has number and gender features only. The latter can be used only predicatively.

<sup>4</sup> This in principle semantic category has morphologic realizations in Russian, unlike in English, and that is why it is treated under grammatical properties.

- syntactic pattern of the adjective-noun construction
- stable prepositional collocations

### Semantic information

- category of the neighbour noun (concrete vs. abstract)
- taxonomic class of the neighbour noun (e.g. “motion”, “time”, “sound”, “colour”, “place”, “emotions”, “natural phenomena”, “hair”, “animals”, “plants”, “texts”, “food and drinks”, “relatives”, “professions”, “stuff”)
- mereological information (e.g. “parts of the body”, “quanta & portions of stuff”)
- topological information (e.g. “containers”, “horizontal spaces”)

Below we will illustrate how each type of information mentioned above is relevant for the filters (for each adjective below, not *all* the rules are given but those that illustrate the importance of the types of information discussed; in other words, we provide a fragment of the filter set for each word).

### 3.1 Grammatical Form of the Adjective: “Long” vs. “Short”

According to standard Russian dictionaries [15] and [16], the adjective *celyj* has two meanings: (a) ‘whole, entire’ and (b) ‘safe, intact’. The corpus data reveal that in the so-called “short” form the adjective is not ambiguous and has the only meaning (b) ‘safe, intact’. Thus, we can formulate the following filter:

*celyj* & short form  $\Rightarrow$  *celyj* ‘safe, intact’ (cf. *cel*, *cela*, *celo*, *cely*)

The resulting disambiguated tag is shown in the table below:

target word	conditions	WSD
<i>celyj</i>	short form	SEM=category: “qualitative”, taxonomy: “physical property”

### 3.2 Grammatical Form of the Noun

The lexico-grammatical category of animacy crucially affects the meaning of attributive adjective construction. Consider the example of the adjective *tolstyj*, which has the meaning ‘fat’ when co-occurring with an animate noun (cf. *tolstyj čelovek* ‘fat man’), and the meaning ‘thick’ in connection with a non-animate noun (cf. *tolstaja kniga* ‘thick book’):

target word	conditions	WSD
<i>tolstyj</i>	+ “animate”	SEM2=category: “qualitative”, taxonomy: “appearance”, evaluation: “negative”
<i>tolstyj</i>	+ “non-animate”	SEM=category: “qualitative”, taxonomy: “size: maximal”

The adjective *pokojnij* ‘deceased/comfortable’ provides a further example for the relevance of animacy for sense disambiguation (cf. *pokojnij otec* ‘deceased father’ vs. *pokojnij divan* ‘comfortable sofa’):

target word	conditions	WSD
<i>pokojnij</i>	+ “animate”	SEM2=category: “relational”
<i>pokojnij</i>	+ “non-animate”	SEM=category: “qualitative”

Another grammatical parameter used in the context filters is the case of the neighbour noun. This kind of rules relies on the fact that some Russian adjectives govern certain cases (cf. also 3.3 below). Interestingly, there is a strong tendency in our data that polysemous adjectives show this property only in one of their meanings, thus, the corresponding uses can be filtered out. The adjective *polnij* can exemplify this type of rule. There are four different senses associated with this word, cf. *vanna polnaja šampanskogo* ‘full tub of Champaign’, *polnij čelovek* ‘fat person’, *opisat’ polnij krug* ‘make a complete cycle’, *polnij durak* ‘absolute idiot’. However, if *polnij* governs genitive it can convey only the ‘full of’ meaning. The filter is applied to all genitive noun constructions apart from those which include the genitive form of the target adjective, since two genitive forms may stand in a syntactic relation of agreement, not of government, and then the target adjective may express any of four relevant senses.

target word	conditions	WSD
<i>polnij</i>	!GEN + S&GEN	SEM=category: “qualitative”, taxonomy: “content”

### 3.3 Syntactic Pattern

Almost all filters take into account the syntactic relation between the target adjective and the neighbour noun. Basic syntactic patterns used for filters are summarized below (A stands for adjective, S for noun, PR for preposition):

- (1) A + S (agreement), cf. *slabij veter* (‘weak’ + ‘wind’: ‘light wind’);
- (2) A + S (government without preposition), cf. *slabij glazami* (‘weak’ + ‘eye’-INSTR: ‘weak-eyed’);
- (3) A + PR + S (government with preposition), cf. *slabij na golovu* (‘weak’ + ‘on’ + ‘head’-ACC: ‘thin in the upper crust’);
- (4) PR + A + S (prepositional phrase), cf. *v slaboj stepeni* (‘in’ + ‘weak’-PRAEP + ‘degree’-PRAEP: ‘to a small degree’).

In most cases, the syntactic information is used alongside with the conditions on morphological and semantic properties of the context. However, sometimes the syntactic pattern alone is sufficient to disambiguate the target adjective. This is mainly the case with the patterns (2) and (3).

Filters that match the pattern (2) operate with the case of the governed noun; an example of such a rule was discussed above (see 3.2, the word *polnij*). Pattern (3) is

applied to those adjectives which in some contexts require the use of a certain preposition. For such adjectives, the preposition has disambiguating power – it signals that the target word can be interpreted unambiguously. The adjective *gluhoj* illustrates this kind of rules. Among its meanings are ‘deaf’ (cf. *gluhoj mal’čik* ‘deaf boy’), ‘dull’ (of sound, cf. *gluhoj zvuk* ‘dull sound’), ‘lonely/remote’ (cf. *gluhaja derevnja* ‘lonely/remote village’), ‘impervious’ (cf. *gluh k dovodam* ‘impervious to argument’). Data analysis reveals that only the latter sense can be conveyed by the construction with the preposition *k* ‘to’:

target word	conditions	WSD
<i>gluhoj</i>	+ <i>k</i>	SEM2= category: “qualitative”, taxonomy: “human property”, evaluation: “negative”

### 3.4 Stable Collocations

When used in stable collocations, words often exhibit specific semantic properties. Such expressions are treated as a special kind of disambiguating contexts. They are represented as syntactic patterns with fixed lexical items. For instance, consider the adjective *černyj* ‘black’ in the collocations like *černyj hod* ‘back entrance’, *černaja magija* ‘black magic’, *na černyj den’* ‘for a rainy day’, etc.:

target word	conditions	WSD
<i>černyj</i>	+ <i>hod</i>	“multiword expression”, SEM2=category: “relational”, taxonomy: “species”
<i>černyj</i>	+ <i>magija</i>	“multiword expression”, SEM2=category: “relational”, taxonomy: “species”
<i>černyj</i>	<i>na + .. + den’</i>	“multiword expression”, SEM2=category: “qualitative”, evaluation: “negative”

### 3.5 Category of the Neighbour Noun

This parameter specifies whether the adjacent noun is concrete or abstract. The relevance of this division for the disambiguation task can be evidenced, e.g., by the adjective *legkij*, which means ‘light’ (of weight) when occurring with a concrete noun, and ‘easy’ or ‘faint’ when referring to an abstract noun (further differentiation is possible based on the taxonomic class of the abstract noun).

target word	conditions	WSD
<i>legkij</i>	+ “concrete”	SEM= category: “qualitative”, taxonomy: “physical property: weight”

### 3.6 Taxonomic Class of the Noun

Taxonomy is undoubtedly the most frequently used parameter to discriminate between the different senses of an adjective. The word *golyj* provides an example of

taxonomic differentiation within the domain of concrete nouns, whereas *holodnyj* illustrates the same issue with abstract nouns.

The adjective *golyj* has among its meanings ‘nude’, cf. *golyj čelovek* ‘nude person’, ‘bare’, cf. *na golom polu* ‘on the bare floor’, and ‘pure’, cf. *golyj spirt* ‘pure alcohol’, each of them imposing restrictions on the taxonomic affiliation of the following noun:

target word	conditions	WSD
<i>golyj</i>	+ “human”	SEM=category: “qualitative”, taxonomy: “physical state”
<i>golyj</i>	+ “space”	SEM2=category: “qualitative”, taxonomy: “appearance”
<i>golyj</i>	+ “stuff”	SEM2=category: “qualitative”, taxonomy: “physical property”

The word *holodnyj* occurs, among others, in the following senses: ‘cold’, cf. *holodnyj veter* ‘cold wind’, ‘cold (of colour)’, cf. *holodnye cveta* ‘cold colours’, and ‘cold/stony’, cf. *holodnyj vzgljad* ‘cold look’. Within the domain of abstract nouns we can draw the following contextual distinctions:

target word	conditions	WSD
<i>holodnyj</i>	+ “weather” + “time”	SEM=category: “qualitative”, taxonomy: “physical property: temperature”
<i>holodnyj</i>	+ “colour”	SEM2=category: “qualitative”, taxonomy: “physical property: colour”
<i>holodnyj</i>	+ “mental sphere” + “emotions” + “psychological states” + “human qualities” + “human behaviour”	SEM2=category: “qualitative”, taxonomy: “human property”

### 3.7 Mereological Class of the Noun

The parameter of taxonomic class, however efficient, cannot account for all relevant semantic properties of the noun. Thus, the adjective *redkij* can show two different senses when used with nouns of the same taxonomic class, cf. *redkaja trava* ‘sparse grass’ and *redkoe rastenie* ‘rare plant’ (both nouns represent the class “plants”). In such cases, the mereological categorization may prove its usefulness for sense disambiguation. In the example above, for instance, the feature “aggregate”, which is characteristic of the word *trava* ‘grass’, is crucial for distinguishing between the two senses of the adjective *redkij*.

target word	conditions	WSD
<i>redkij</i>	+ “plant” & “aggregate”	SEM=category: “relational”, taxonomy: “physical property”

### 3.8 Topological Class of the Noun

Topology, i.e. geometric features of the object referred to by a noun, is a further parameter which may be of use when the taxonomic classification fails to differentiate between senses. The adjective *tugoj* illustrates the case in question. When used with a



noun of object, it usually means ‘tight’, cf. *tugoj uzel* ‘tight knot’. An exception to this is objects associated to the topological type of “containers”, cf. *tugoj košelek* ‘fat purse’.

target word	conditions	WSD
<i>tugoj</i>	+ “containers”	SEM2=category: “qualitative”, taxonomy: “size:maximal”
<i>tugoj</i>	+ “concrete”	SEM=category: “qualitative”, taxonomy: “physical property”

#### 4 Conclusion: Theoretical Extensions of the Research

There is a total amount of about 300 frequently used polysemous adjectives denoting qualities in RNC (those that occur more than 2000 times per 140 mln words). Presently 240 of them are supplied with disambiguating filters. The results obtained show that the method discussed above is highly efficient in those contexts where an adjective is adjacent to the modified noun. However, non-adjacent uses of adjectives, in particular predicative adjectives, are more problematic for the current version of the rules. In other words, we achieve a high precision rate (93%), but a lower recall rate (47%). As a next step, we plan to develop rules which would account for non-adjacent positions of adjectives.

Due to the method applied, the practical task of WSD may have theoretical extensions, which concern the nature of polysemy and the principles of semantic evolution.

- The procedure of a rule-based approach helps to specify the linguistically relevant classes of nouns for word-sense disambiguation. The classes which have proven to be the most useful for meaning differentiation are ‘animate’ (including ‘human’) vs. ‘non-animate’, and ‘abstract’ vs. ‘concrete’. The change between these classes always leads to a shift in meaning of a modifying adjective. However, the question is, what *other* classes may be of any relevance for changing the meaning of a word. The fact that a taxonomic class is used in filters proves that it is cognitively relevant.
- The analysis done for the filters makes it possible to identify the regular patterns of semantic shifts in adjective meaning (cf. adjectives with the basic meaning of physical property combined with a noun of the taxonomic class ‘human’ regularly obtain the sense of the non-physical property of a person: *mjagkij divan* ‘soft sofa’– *mjagkij čelovek* ‘tender person’, *legkij čemodan* ‘light bag’– *legkij čelovek* ‘easy person to get along with’).
- The important characteristic of a semantic shift is not only the initial and the final meanings of an adjective, but the nature of the shift itself. The two main types of shifts are metaphor and metonymy. We have seen some examples of metaphorical shifts above; regular metonymy can be instantiated by the application of a human property to body parts (cf. *dobryj čelovek* ‘kind person’ – *dobrye glaza* ‘kind eyes’). Russian adjectives offer a fertile ground for research on the so far less studied domain of metonymy occurring through parts of speech (cf. [17]): many adjectives are metonymically associated with the adverbs derived from them (cf.

*redkij* ‘thin-growing, sparse’: *redkaja boroda* ‘thin beard’ – *derev’ya redko rastut* ‘the trees grow sparsely’). The cataloguing of the semantic shifts of adjectives can contribute to the general theory of metaphor and metonymy and can extend the list of known metaphoric and metonymic patterns (cf. [18]).

- The reasons for the dissimilar behavior of synonyms and antonyms should be thoroughly examined. Although synonymic and antonymic pairs sometimes reveal similar meaning shifts (cf. *sil’ny čelovek* – *sil’ny harakter* – *sil’ny učeník* ‘strong man’ – ‘strong character’ – ‘strong pupil’, and *slaby čelovek* – *slaby harakter* – *slaby učeník* – ‘weak man’ – ‘weak character’ – ‘weak pupil’), they are not always symmetric: consider, for example, the antonymic adjectives *dikij* ‘wild’ and *domašnij* ‘domestic’. The first of them, beside its main meaning ‘living in the wild, not cultivated’, develops such senses as the human property ‘strange’ (*dikiy rebenok* ‘strange child’), negative ‘odd, strange’ (*dikaja vyhodka* ‘odd action’), and high degree ‘wild’ (*dikij vostorg* ‘wild gaiety’). The antonym of *dikij* – the adjective *domašnij* – has only the meaning that corresponds to the idea of house (*domašn’aja rabota* ‘homework’, *domašnije tapočki* ‘house slippers’ and so on.
- Gaining an evaluational polarity (positive / negative) or a change in polarity constitutes a further type of adjective meaning shifts. Several constraints on possible transitions can be observed. Thus, for instance, we have not encountered so far any case of an adjective that has changed from negative to positive polarity.

So the practical problem of word sense disambiguation turns to be a challenge to theoretical semantics and lexicology. The more language data is involved in the analysis, the better we can observe the systematic character of the lexicon organization and the regularity of the models of semantic evolution.

## Acknowledgments

The present research is supported by RFBR foundation (ref. 08-06-00197). We would like to thank two anonymous reviewers for their constructive comments that helped to improve the manuscript.

## References

1. Agirre, E., Edmonds, P.: Introduction. In: Agirre, E., Edmonds, P. (eds.) *Word Sense Disambiguation: Algorithms and Applications*, pp. 1–28. Springer, Heidelberg (2006)
2. Fillmore, C.J., Kay, P., O’Connor, K.T.: Regularity and Idiomaticity in Grammatical Constructions: The Case of LET ALONE. *Language* 64, 501–538 (1988)
3. Goldberg, A.E.: *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago University Press, Chicago (1995)
4. Jackendoff, R.: *Twistin’ the night away*. *Language* 73, 534–559 (1997)
5. Jackendoff, R.: *Semantic Structures*. MIT Press, Cambridge (1990)
6. Pustejovsky, J.: Type Coercion and Lexical Selection. In: Pustejovsky, J. (ed.) *Semantics and the Lexicon*, pp. 73–94. Kluwer Academic Publishers, Dordrecht (1993)
7. Partee, B.H.: Compositionality and Coercion in Semantics: The Dynamics of Adjective Meaning. In: Bouma, G., Kraemer, I., Zwarts, J. (eds.) *Cognitive Foundations of Interpretation*, pp. 145–161. Royal Netherlands Academy of Arts and Sciences, Amsterdam (2007)

8. Lashevskaja, O.: Corpus-aided Construction Grammar: Semantic Tools in the Russian National Corpus. In: Proceedings of the Second International Conference of the German Cognitive Linguistics Association, Munich, October 5-7 (2006)
9. Apresjan, Y.D., Boguslavskij, I.M., Iomdin, B.L., Iomdin, L.L., Sannikov, A.V., Sannikov, V.Z., Sizov, V.G., Tsinman, L.L.: Sintaksicheski i semanticheski annotirovannyj korpus russkogo jazyka: sovremennoe sostojanie i perspektivy (in Russian). In: Natsionalnyj korpus russkogo jazyka: 2003—2005, pp. 193–214. Indrik, Moscow (2003)
10. Rakhilina, E.V., Kobritsov, B.P., Kustova, G.I., Shemanaeva, O.Y.: Mnogoznachnost' kak prikladnaja problema: Leksiko-semanticheskaja razmetka v NKRJa (in Russian). In: Laufer, N.I., Narinjani, A.S., Selegej, V.P. (eds.) *Komputernaja lingvistika i intelektual'nye tehnologii: Trudy mezhdunarodnoj konferencii «Dialog 2006»*, pp. 445–450. Izd-vo RGGU, Moscow (2006)
11. Rakhilina, E.V.: Kognitivnyj analiz predmetnyx imen: semantika i sočetaemost' (in Russian). *Russkie slovari*, Moscow (2000)
12. Paduceva, E.V.: Dinamičeskie modeli v semantike leksiki (in Russian). *Jazyki slavjanskoj kul'tury*, Moscow (2004)
13. Kustova, G.I.: Tipy proizvodnyx značenij i mexanizmy jazykovogo rasširenija (in Russian). *Jazyki slavjanskoj kul'tury*, Moscow (2004)
14. Talmy, L.: How language structures space. In: Pick, H., Acredolo, L. (eds.) *Spatial Orientation: Theory, Research, and Application*, pp. 225–282. Plenum Press, New York (1983)
15. Ozhegov, S.I., Shvedova, N.J.: *Tolkovyj slovar' russkogo jazyka* (in Russian). *Russkij jazyk*, Moscow (1992)
16. Evgen'eva, A.P. (ed.): *Slovar' russkogo jazyka* (in Russian), vol. 1-4. *Russkij jazyk*, Moscow (1999)
17. Radden, G., Kövecses, Z.: Towards a Theory of Metonymy. In: Panther, K.-U., Radden, G. (eds.) *Metonymy in language and thought*, pp. 17–59. Benjamins, Amsterdam (1999)
18. Lakoff, G.: *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. University of Chicago Press, Chicago (1987)